Covariance estimation in terms of Stokes parameters with application to vector sensor imaging

Ryan Volz*, Mary Knapp†, Frank D. Lind*, Frank C. Robey‡
Massachusetts Institute of Technology
*Haystack Observatory, Westford, MA
†Department of Earth, Atmospheric, and Planetary Science, Cambridge, MA
‡Lincoln Laboratory, Lexington, MA

***Abstract*—Vector sensor imaging presents a challenging problem in covariance estimation when allowing arbitrarily polarized sources. We propose a Stokes parameter representation of the source covariance matrix which is both qualitatively and computationally convenient. Using this formulation, we adapt the proximal gradient and expectation maximization (EM) algorithms and apply them in multiple variants to the maximum likelihood and least squares problems. We also show how EM can be cast as gradient descent on the Riemannian manifold of positive definite matrices, enabling a new accelerated EM algorithm. Finally, we demonstrate the benefits of the proximal gradient approach through comparison of convergence results from simulated data.**

I. Introduction

A. Vector sensor imaging

The developments of this paper are motivated by the electromagnetic vector sensor imaging problem: estimating the magnitude, polarization, and direction of plane wave sources from a sample covariance matrix of vector measurements. A vector sensor (example shown in Figure 1) measures the electromagnetic field at a single point using three orthogonal dipole elements and three orthogonal loop elements with a common phase center, producing a six-element measurement vector [2]. This vector represents the full state of the electromagnetic field at the sensor's location [3], implying sensitivity from all directions and complete polarization information. These properties make even a single vector sensor ideal for determining the direction of arrival of signal sources [2, 4, 5], nulling or isolating specific sources, and maximizing the statistics collected at a single location [3]. In this application, we seek to use a single vector sensor to make a map of signal strength and polarization as a function of direction of arrival.

We model the measured signals in terms of a collection of point sources distributed equally in angle on the surrounding sphere as in [3]. The magnitude and phase of each source at time sample n of N is collected in the complex vector c_n. The six-element measurement vector r_n is then given by

$$r_n = Ac_n + w_n, \tag{1}$$

where A is the matrix of steering vectors describing the contribution from each individual source and w_n is measurement noise. We assume that the sources have an arbitrary phase with respect to time over the N samples, so we impose a probabilistic model where the source and noise vectors are zero-mean complex normal with covariances Σ and $\sigma\mathrm{I}$:

$$\begin{aligned} c_n &\sim \mathcal{CN}(0,\, \Sigma) \quad \forall n \\ w_n &\sim \mathcal{CN}(0,\, \sigma\mathrm{I}) \quad \forall n \\ r_n &\sim \mathcal{CN}(0,\, A\Sigma A^* + \sigma\mathrm{I}) \quad \forall n, \end{aligned} \tag{2}$$

where I denotes the identity matrix. Assuming a known estimate of the noise variance σ, our objective is to estimate the source covariance Σ from the measurements r_n.

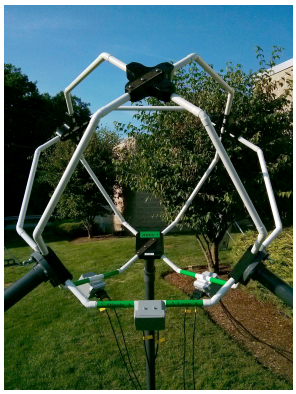

Figure 1. *Atom antenna [1], an electromagnetic vector sensor.* The antenna is composed of three orthogonal loop and dipole elements with a common phase center, measuring the complete electromagnetic field in a six-element vector.

B. Covariance estimation

One approach to estimating the source covariance is to seek the maximum likelihood solution for Σ, with the likelihood given by the complex normal distribution for r_n.

Distribution A: approved for public release; unlimited distributon. This material is based upon work supported by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Assistant Secretary of Defense for Research and Engineering.

Equivalently, we can minimize the negative log-likelihood, leading to the objective function

$$H_{ml}(\Sigma) = \log\det(A\Sigma A^* + \sigma \mathrm{I}) + \mathrm{tr}((A\Sigma A^* + \sigma \mathrm{I})^{-1} S). \quad (3)$$

The measurements are present only in the sample covariance matrix S, making it a sufficient statistic:

$$S = \frac{1}{N} \sum_{n=0}^{N-1} r_n r_n^*. \quad (4)$$

Another approach, perhaps more convenient because of its convexity and relative simplicity, is to minimize the squared error between the estimated measurement covariance and the sample covariance. This gives the least-squares objective

$$H_{ls}(\Sigma) = \tfrac{1}{2} \| S - (A\Sigma A^* + \sigma \mathrm{I}) \|_2^2. \quad (5)$$

For both cases, we seek to solve an optimization problem of the form

$$\begin{aligned} \underset{\Sigma}{\text{minimize}} \quad & H(\Sigma) \\ \text{subject to} \quad & \Sigma \succeq 0, \end{aligned} \qquad (\mathrm{P_{ml}}, \mathrm{P_{ls}})$$

which minimizes the given objective while enforcing the constraint that the covariance be positive semidefinite (PSD). The choice of objective function can depend on a number of factors: convergence guarantees, algorithm performance, and correctness. Maximum likelihood solves the desired problem from a statistical standpoint, but it may be that least-squares also produces an acceptable solution for a given case while being more practical to implement.

II. Source signal polarization

A. Independent polarized source model

It is possible to solve the maximum likelihood and least-squares covariance estimation problems for a general Σ, but a significant simplification arises by assuming that the sources are statistically independent for each direction in the sky. This does not, however, imply that the source covariance matrix Σ is diagonal. In order to allow for sources with arbitrary polarization, each source must be represented in a basis of two orthogonal polarizations. The Jones vector z fully quantifies the state of a two-dimensional transverse wave at a given location and time:

$$z = \begin{bmatrix} E_h e^{i\phi_h} \\ E_v e^{i\phi_v} \end{bmatrix} = \begin{bmatrix} z_h \\ z_v \end{bmatrix}. \quad (6)$$

Each component z_a of the Jones vector gives the wave's amplitude and phase as a complex number for the corresponding basis direction, in this case horizontal and vertical. Therefore, each independent arbitrarily-polarized source requires two entries in c_n for its Jones vector components.

Representing the horizontal and vertical basis components of each source with corresponding entries in vectors h_n and v_n, we can partition the measurement equation (1) as

$$r_n = \begin{bmatrix} A_h & A_v \end{bmatrix} \begin{bmatrix} h_n \\ v_n \end{bmatrix} + w_n, \quad (7)$$

where A_h and A_v give the steering vectors for horizontal and vertical polarizations at each source location. This leads to a structured source covariance,

$$\Sigma = \mathbf{E}\left(\begin{bmatrix} h_n \\ v_n \end{bmatrix} \begin{bmatrix} h_n^* & v_n^* \end{bmatrix} \right) = \begin{bmatrix} \Sigma_{hh} & \Sigma_{hv} \\ \Sigma_{hv}^* & \Sigma_{vv} \end{bmatrix}. \quad (8)$$

Note that assuming independent sources for each pointing direction implies that each of the blocks Σ_{hh}, Σ_{hv}, and Σ_{vv} must be diagonal.

B. Stokes parameter formulation

It is convenient both qualitatively and computationally to represent Σ in terms of Stokes parameters:

$$\Sigma = \tfrac{1}{2} \begin{bmatrix} \mathrm{diag}(I+Q) & \mathrm{diag}(U-iV) \\ \mathrm{diag}(U+iV) & \mathrm{diag}(I-Q) \end{bmatrix}, \quad (9)$$

where I, Q, U, and V are vectors with corresponding entries for each independent source. This satisfies the typical definition of Stokes parameters in terms of second-order polarization moments:

$$\begin{aligned} I &= \mathrm{diag}(\Sigma_{hh} + \Sigma_{vv}) & U &= 2\Re(\mathrm{diag}(\Sigma_{hv})) \\ Q &= \mathrm{diag}(\Sigma_{hh} - \Sigma_{vv}) & V &= -2\Im(\mathrm{diag}(\Sigma_{hv})). \end{aligned} \quad (10)$$

I can be interpreted as total intensity, while Q, U, and V describe different modes of polarization.

C. Projection onto the positive semidefinite cone

Parametrization in terms of Stokes parameters is convenient because it becomes easy to project a covariance $\hat{\Sigma}$ onto the positive semidefinite cone:

$$\begin{aligned} \underset{\Sigma}{\arg\min} \quad & \left\| \Sigma - \hat{\Sigma} \right\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \Sigma \succeq 0. \end{aligned} \qquad (\mathrm{P_{proj}})$$

With $\hat{\Sigma}$ represented by Stokes parameters $\hat{I}$, $\hat{Q}$, $\hat{U}$, and $\hat{V}$, the Frobenius norm in ($\mathrm{P_{proj}}$) becomes

$$\tfrac{1}{2} \left\| I - \hat{I} \right\|_2^2 + \tfrac{1}{2} \left\| Q - \hat{Q} \right\|_2^2 + \tfrac{1}{2} \left\| U - \hat{U} \right\|_2^2 + \tfrac{1}{2} \left\| V - \hat{V} \right\|_2^2. \quad (11)$$

Similarly, the positive semidefinite condition becomes

$$\Sigma \succeq 0 \Leftrightarrow \begin{array}{c} I \succeq 0 \\ I^2 \succeq Q^2 + U^2 + V^2, \end{array} \quad (12)$$

where $\succeq$ on the right-hand side implies elementwise comparison.

This vector optimization problem is separable into scalar components, so it suffices to solve the problem for each (I_k, Q_k, U_k, V_k) tuple and apply the resulting solution element-wise to the vector Stokes parameters. First define

$$P_k = \sqrt{\hat{Q}_k^2 + \hat{U}_k^2 + \hat{V}_k^2}. \quad (13)$$

If $\hat{I}_k \geq P_k$, no projection is needed and the solution is

$$I_k = \hat{I}_k \quad Q_k = \hat{Q}_k \quad U_k = \hat{U}_k \quad V_k = \hat{V}_k. \quad (14)$$

Otherwise, evaluating the KKT conditions [6] gives a closed-form expression for the unique minimizer:

$$I_k = \max\left(\tfrac{1}{2}\left(\hat{I}_k + P_k\right),\, 0\right)$$
$$Q_k = \frac{I_k}{P_k}\hat{Q}_k \qquad U_k = \frac{I_k}{P_k}\hat{U}_k \qquad V_k = \frac{I_k}{P_k}\hat{V}_k. \tag{15}$$

Essentially, non-physical Stokes components with the total intensity less than the polarized intensity are scaled to be fully polarized. Since the general solution to ($\mathrm{P}_{\mathrm{proj}}$) requires an eigendecomposition [7], it is clear that projection in terms of Stokes parameters produces a significant simplification.

III. Algorithms

A. Expectation maximization

Application of the expectation maximization (EM) algorithm [8] to the maximum likelihood covariance estimation problem is well-known [9, 10], and its formulation is given in Algorithm 1. It is an iterative algorithm that makes use

Algorithm 1 EM for (P_{ml}) [9]

given Σ^0, measurements $(S,\, \sigma)$
repeat
$\Sigma^{k+1} \leftarrow \Sigma^k - \Sigma^k\, \mathbf{grad}_{H_{ml}}(\Sigma^k)\Sigma^k$
until stopping criterion is satisfied

of the gradient of the negative log-likelihood function,

$$\mathbf{grad}_{H_{ml}}(\Sigma) = A^* R^{-1}(R - S)R^{-1}A, \tag{16}$$

where

$$R = A\Sigma A^* + \sigma \mathrm{I}. \tag{17}$$

A convenient property of EM is that it implicitly enforces the condition that Σ be positive semidefinite.

B. Proximal gradient

A similar algorithm that is less well-known with respect to the covariance estimation problem is proximal gradient [11], given in Algorithm 2 for a general setting. It solves

Algorithm 2 Proximal Gradient (PG) [11]

given step size μ, x^0
repeat
$z^{k+1} \leftarrow x^k - \mu\, \mathbf{grad}_H(x^k)$
$x^{k+1} \leftarrow \mathbf{prox}_{\mu F}(z^{k+1})$
until stopping criterion is satisfied

split-objective problems of the form

$$\underset{x}{\text{minimize}} \quad F(x) + H(x), \tag{$\mathrm{P}_{\mathrm{proxgrad}}$}$$

where F must admit a proximal operator,

$$\mathbf{prox}_F(v) = \underset{x}{\arg\min}\, \left(F(x) + \tfrac{1}{2}\|x - v\|_2^2\right),$$

and H must be smooth with a gradient given by $\mathbf{grad}_H(x)$.

For both the maximum likelihood and least-squares covariance estimation problems, we already have a smooth H. In the least-squares case,

$$\mathbf{grad}_{H_{ls}}(\Sigma) = A^*(A\Sigma A^* + \sigma \mathrm{I} - S)A. \tag{18}$$

The positive semidefiniteness constraint can be incorporated into F as an indicator function, making the proximal operator into Euclidean projection onto the PSD cone.

Though this projection adds a wrinkle in comparison to the EM algorithm, proximal gradient can incorporate improvements such as adaptive step size [12, 13] and acceleration [14, 15] that enable improved convergence. In addition, it provides a straightforward path to include additional constraints or regularization that can be useful for solving variants of the covariance estimation problem.

C. EM as Riemannian proximal gradient

Given the similarities between the EM and proximal gradient algorithms, it is not surprising that they share a connection when the problem is viewed through the lens of gradient descent on the manifold of positive definite matrices. Manifold optimization relies on the Riemannian gradient, which provides the direction in the manifold's tangent plane along which the function increases most [16, 17]. On the $\Sigma \succ 0$ manifold, the Riemannian gradient of the maximum likelihood objective is a modified Euclidean gradient [18]:

$$\mathbf{grad}^{\mathrm{PD}}_{H_{ml}}(\Sigma) = \Sigma\, \mathbf{grad}_{H_{ml}}(\Sigma)\Sigma. \tag{19}$$

Notice that the Riemannian gradient is precisely the expression that appears in the EM step of Algorithm 1. From this perspective, EM is simply unit-step gradient descent on the positive definite manifold.

Introduction of a variable step size to EM is possible through an appeal to the manifold optimization concept of a retraction. Useful in this case is the projective retraction, which takes a Euclidean step along the Riemannian gradient followed by a Euclidean projection onto the manifold [17]:

$$\begin{aligned}\Sigma^{k+1} &= \mathrm{Ret}_{\mathrm{proj}}(\Sigma^k,\, \mu\, \mathbf{grad}^{\mathrm{PD}}_{H_{ml}}(\Sigma^k)) \\ &= \mathbf{prox}_{I_{\Sigma\succ 0}}(\Sigma^k - \mu\, \mathbf{grad}^{\mathrm{PD}}_{H_{ml}}(\Sigma^k)).\end{aligned} \tag{20}$$

Unfortunately, projection onto the open set of positive definite matrices is not well-defined; we can approximate by projecting onto the positive semidefinite cone instead:

$$\Sigma^{k+1} = \mathbf{prox}_{I_{\Sigma\succeq 0}}(\Sigma^k - \mu\, \mathbf{grad}^{\mathrm{PD}}_{H_{ml}}(\Sigma^k)). \tag{21}$$

This looks like proximal gradient (Algorithm 2) applied to the maximum likelihood problem, only with the Riemannian gradient replacing the Euclidean gradient.

In other words, by taking proximal gradient and using a Riemannian gradient step, one can produce a new family of "EM" algorithms. Because we can then apply convergence improvements developed for proximal gradient, this new Riemannian proximal gradient algorithm may provide significant benefit over EM.

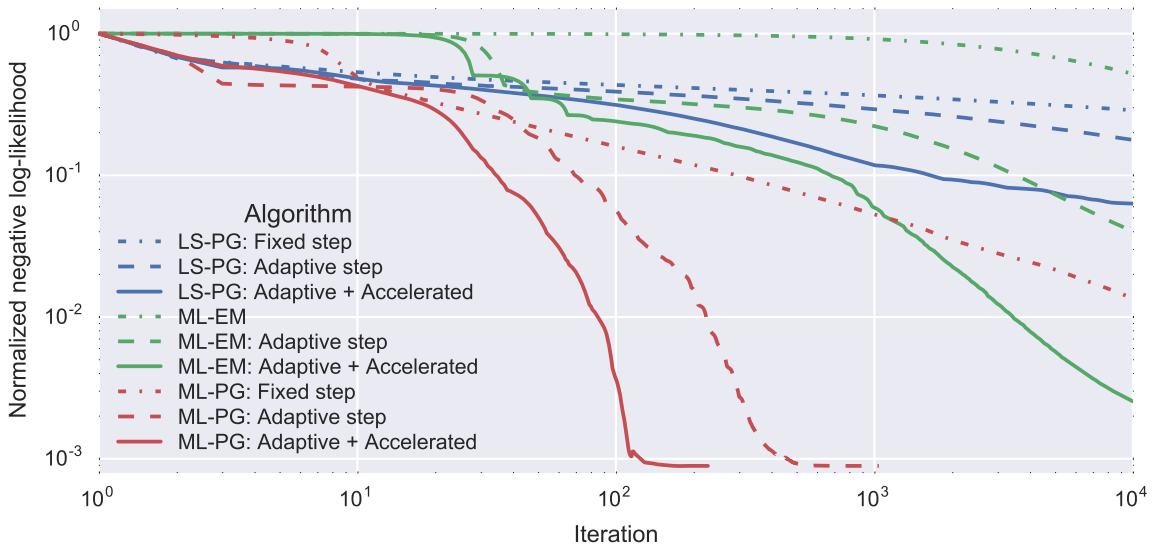

Figure 2. *Convergence in likelihood for a random problem instance.* The accelerated, adaptive-step variant converges in the fewest iterations for all problem formulations: least-squares with proximal gradient (LS-PG), maximum likelihood with EM or Riemannian proximal gradient (ML-EM), and maximum likelihood with proximal gradient (ML-PG).

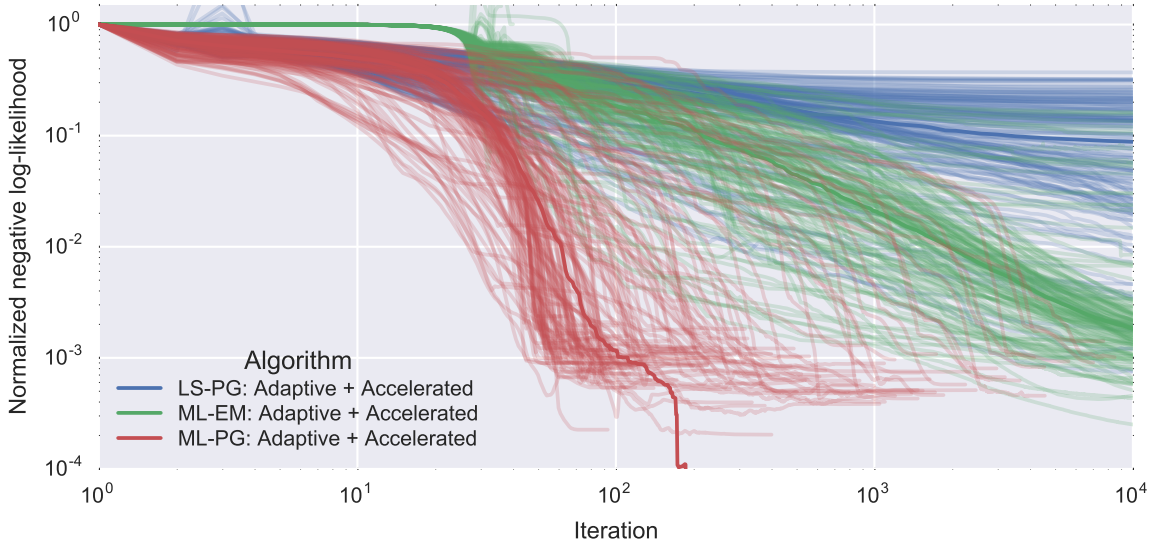

Figure 3. *Convergence in likelihood over all runs for the fastest variants.* Trends from the single run case hold over many instances. ML-PG readily converges to a sparse solution, so it is the easy choice for vector sensor imaging. ML-EM is relatively slow because it is constrained to the positive definite manifold. LS-PG is not competitive enough to justify the relative simplicity of its implementation.

IV. Simulation convergence comparisons

To compare algorithm convergence rates, we simulated 150 instances of vector sensor measurements for two randomly located and polarized point sources. Figure 2 shows the normalized negative log-likelihood, scaled and shifted to start at one and converge to zero, for a single problem instance as a function of iteration number. It gives results for proximal gradient applied to the maximum likelihood (ML-PG) and least squares (LS-PG) problems and EM or Riemannian proximal gradient applied to the maximum likelihood problem (ML-EM). Each algorithm was tested in variants with either a fixed step, adaptive step, or adaptive step with acceleration. As expected, the accelerated, adaptive-step variant converges in the fewest iterations for all problem formulations. The improvements are significant in this instance, with fixed step ML-PG converging in over 10,000 iterations while the accelerated adaptive-step variant converges in about 200 iterations.

To best compare the problem formulations and algorithms between each other, Figure 3 shows the convergence of all 150 instances for the three configurations using an adaptive step size and acceleration. Fastest convergence in general is achieved when solving the maximum likelihood problem. Between standard proximal gradient using the

Euclidean gradient and our "EM" proximal gradient using the Riemannian gradient, standard proximal gradient generally converges faster. The least-squares formulation, even though it does eventually converge to the maximum likelihood solution in these cases, is not competitive enough to justify the relative simplicity of its implementation.

The main performance difference between formulations likely arises from how well they handle the sparsity of the true solution. Since only two sources exist in the true solution, the source covariance is mostly zero and decidedly positive semidefinite. Since EM, with the Riemannian gradient, operates on the positive definite manifold, it is at a decided disadvantage in trying to converge to a positive semidefinite solution. It is for this reason that we find ML-PG most suitable for use with vector sensor imaging. However, the new ML-EM algorithm may be of great benefit in other problem settings where classic EM already excels and a positive definite solution is expected.

V. Conclusion

Vector sensor imaging poses a challenge for covariance estimation. The EM algorithm solves the maximum likelihood problem, but its slow rate of convergence makes EM less than ideal. The proximal gradient algorithm, in contrast, is well suited to the vector sensor maximum likelihood problem and can also be applied to the least-squares covariance estimation problem. EM and proximal gradient are similar, but EM employs the Riemannian gradient over the manifold of positive definite matrices. From this, we formulated a new Riemannian proximal gradient algorithm combining the Riemannian gradient of EM with the projection and variable step size of proximal gradient.

The difficulty in applying any form of proximal gradient is that it employs projection onto the positive semidefinite cone, which can be slow and pose numerical problems for estimating a general covariance matrix. We showed that for independent but arbitrarily polarized signal sources, Stokes parameters can be used to represent the source covariance matrix and simplify projection onto the PSD cone.

We compared the convergence of the proximal gradient algorithms applied to both the maximum likelihood and least-squares problems using simulated data. For the vector sensor imaging problem, we showed that the Euclidean proximal gradient maximum likelihood algorithm converged in the fewest iterations, likely because the solution was sparse and unreachable from exact Riemannian gradient steps. Nevertheless, we also showed that Riemannian proximal gradient improves upon EM, so it may be useful in other applications for which EM is better-suited.

References

[1] M. Knapp, R. Volz, F. D. Lind, F. C. Robey, A. Fenn, K. Johnson, M. Silver, A. Morris, and S. Klein, "HF Vector Sensor for Radio Astronomy: Ground Testing Results," in *AIAA SPACE 2016*, ser. AIAA SPACE Forum, American Institute of Aeronautics and Astronautics, Sep. 2016. DOI: 10.2514/6.2016-5498.

[2] A. Nehorai and E. Paldi, "Vector-sensor array processing for electromagnetic source localization," *IEEE Transactions on Signal Processing*, vol. 42, no. 2, pp. 376–398, Feb. 1994, ISSN: 1053587X. DOI: 10.1109/78.275610.

[3] M. Knapp, F. C. Robey, R. Volz, F. D. Lind, A. Fenn, A. Morris, M. Silver, S. Klein, and S. Seager, "Vector antenna and maximum likelihood imaging for radio astronomy," in *2016 IEEE Aerospace Conference*, Mar. 2016, pp. 1–17. DOI: 10.1109/AERO.2016.7500688.

[4] K.-C. Ho, K.-C. Tan, and B. T. G. Tan, "Estimation of directions-of-arrival of partially polarized signals with electromagnetic vector sensors," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 5, May 1996, 2900–2903 vol. 5. DOI: 10.1109/ICASSP.1996.550160.

[5] K. Wong and M. Zoltowski, "Uni-vector-sensor ESPRIT for multisource azimuth, elevation, and polarization estimation," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 10, pp. 1467–1474, Oct. 1997, ISSN: 0018-926X. DOI: 10.1109/8.633852.

[6] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK ; New York: Cambridge University Press, 2004, ISBN: 0521833787.

[7] N. J. Higham, "Computing a nearest symmetric positive semidefinite matrix," *Linear Algebra and its Applications*, vol. 103, pp. 103–118, May 1988, ISSN: 0024-3795. DOI: 10.1016/0024-3795(88)90223-6.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977, ISSN: 0035-9246.

[9] F. C. Robey, "A Covariance Modeling Approach to Adaptive Beamforming and Detection," Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, Technical Report 918, Jul. 1991, p. 137.

[10] A. D. Lanterman, "Statistical Imaging in Radio Astronomy via an Expectation-Maximization Algorithm for Structured Covariance Estimation," in *Statistical Methods in Imaging: IN Medicine, Optics, and Communication*, J. A. O'Sullivan, Ed., Springer-Verlag, to appear. [Online]. Available: http://users.ece.gatech.edu/lanterma/papers/dlsradio.pdf (visited on 03/25/2015).

[11] N. Parikh and S. P. Boyd, "Proximal Algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, Jan. 2014. DOI: 10.1561/2400000003.

[12] K. Scheinberg, D. Goldfarb, and X. Bai, "Fast First-Order Methods for Composite Convex Optimization with Backtracking," *Foundations of Computational Mathematics*, vol. 14, no. 3, pp. 389–417, Mar. 2014. DOI: 10.1007/s10208-014-9189-9.

[13] S. R. Becker, E. J. Candès, and M. C. Grant, "Templates for convex cone problems with applications to sparse signal recovery," en, *Mathematical Programming Computation*, vol. 3, no. 3, pp. 165–218, Aug. 2011, ISSN: 1867-2957. DOI: 10.1007/mpc.v3i3.58.

[14] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009. DOI: 10.1137/080716542.

[15] B. O'Donoghue and E. J. Candès, "Adaptive Restart for Accelerated Gradient Schemes," en, *Foundations of Computational Mathematics*, pp. 1–18, 2013, ISSN: 1615-3375, 1615-3383. DOI: 10.1007/s10208-013-9150-3.

[16] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, English. Princeton, N.J.; Woodstock: Princeton University Press, 2008, ISBN: 978-1-4008-3024-4.

[17] P.-A. Absil and J. Malick, "Projection-like Retractions on Matrix Manifolds," *SIAM Journal on Optimization*, vol. 22, no. 1, pp. 135–158, Jan. 2012, ISSN: 1052-6234. DOI: 10.1137/100802529.

[18] S. Sra and R. Hosseini, "Conic Geometric Optimization on the Manifold of Positive Definite Matrices," *SIAM Journal on Optimization*, vol. 25, no. 1, pp. 713–739, Jan. 2015. DOI: 10.1137/140978168.